BenchmarkXPRT

Exploring the TouchXPRT 2014 Benchmark

TouchXPRT 2014 is a benchmark for evaluating the capabilities of your Windows 8 and Windows RT devices.

May 9, 2014

TOUCHXPRT

TABLE OF CONTENTS

INTRODUCTION

This paper explains the concepts behind TouchXPRT 2014. TouchXPRT 2014 is a tool created to evaluate how well a Windows 8/8.1 or Windows RT laptop or tablet handles common media tasks. It provides an intuitive user interface, a quick runtime for individual and overall tests, and easy-to-understand results. We explain the development guidelines common to all BenchmarkXPRT benchmarks as well as the specific goals and assumptions of TouchXPRT 2014. We also discuss the structure of each of TouchXPRT 2014's component tests in detail. Finally, we demonstrate how the benchmark calculates results, how to share results, and how to participate in the BenchmarkXPRT community.

DEVELOPMENT PROCESS

We use a unique design methodology when developing the BenchmarkXPRT benchmarks. Instead of the closed, bottom-up approach used by many benchmarking efforts, we utilize an open, top-down approach that includes the Development Community. Our approach starts by taking input from the community and examining the most common use cases. We then write a Request for Comment (RFC) proposing use cases to incorporate into the benchmark. Once we have the RFC written, we publish it to the community.

The community's input on the RFC guides the drafting of a design document. The design document drives the implementation of the community preview. After releasing the community preview to the community, we take the input from the preview period and finalize the code to create a general release.

TOUCHXPRT 2014: THE DETAILS

TouchXPRT 2014 shares the common characteristics of the BenchmarkXPRT family of benchmarks:

- It is easy to use.
- It contains relatable workloads.
- It produces easy-to-understand results.

TouchXPRT reports how well a touch-capable device running Windows 8/8.1 or Windows RT handles common media tasks. To do this, the benchmark uses five tests built around user-based scenarios that implement the kinds of photo, video, and MP3 tasks users perform on their device every day. The test measures how long a device takes to complete each of those tasks and produces an overall score. After testing, users can compare their device's overall score to others at TouchXPRT.com. To create an initial list of roles or usage categories, we spent some time looking at what is available on the iTunes App Store, the Google Play store, and the Windows Store. From this, we created an initial set of usages for TouchXPRT workloads:

- Beautify Photos
- Blend Photos
- Convert Videos for Sharing
- Create Music Podcast
- Create Slideshow from Photos

Obviously, this list does not reflect all the ways people use their touch devices, but we believe it accurately reflects the types of tasks that stress devices. When we measure such tasks accurately, they provide a solid basis for identifying performance differences between devices.

UI enhancements

While the UI for TouchXPRT generally got high marks for its appearance, there were a few flaws that people noticed. We addressed these in the new version.

- The Run All button is now always visible, instead of on the Bottom App Bar.
- TouchXPRT 2014 has an integrated results viewer.
- TouchXPRT 2014 can be automated, so that the tests run without human intervention.
- TouchXPRT 2014 has an integrated mechanism for submitting results to be published on the PT Web site. We discuss publishing results in more detail in the Submitting results section below.

The component tests

We updated the tests in TouchXPRT 2014 to ensure that they better reflect the more demanding ways people are using their devices today. We also enhanced the appearance of some of the tests, both to make them more visually appealing and to make it clearer what the test is doing.

Test scenario details

- **Beautify Photos.** This scenario represents a user applying three separate effects to six sets of four photos in a photo album (24 photos total, 16 megapixels). TouchXPRT applies each effect (Sharpen, Spotlight, and Emboss) to each of the photos and measures the elapsed time during each application. The sum of all measured times is reported as the scenario score (in seconds).

  Improvements from TouchXPRT 2013 include the following:
  - The test is now multithreaded to better reflect the performance of multi-core devices.
  - The photos are higher resolution. They are now 16MP instead of 10.2MP.
  - Testing showed that we could reduce the number of photos without compromising the results. So, TouchXPRT reduced the number of photos from 50 to 24, which saves testing time.

- **Blend Photos.** This scenario comprises three tasks, and each task represents a different way of blending photos. The first task, Photo Collages, represents a user creating photo collages by blending three sets of eight photos from a photo album (24 photos total, 16 megapixels) into three separate collages, and adding a watermark to each collage. The second task, Smart Merge, removes moving objects from a burst of photos (5 photos total, 5 megapixels). The third task, HDR, aligns photos and then performs an HDR operation (4 photos total, 6 megapixels). TouchXPRT measures the elapsed time for each of the three blending tasks and reports the sum of all measured times as the scenario score. Blend Photos includes the operations used in the TouchXPRT 2013 Prepare Photos for Sharing test.

  Improvements from TouchXPRT 2013 include the following:
  - There are two new tasks: Smart Merge and HDR.
  - The new scenarios are multithreaded to better reflect the performance of multi-core devices.

- o The photos for collages are higher resolution. They are now 16MP instead of 10.2MP.

- Convert Videos for Sharing. This scenario represents a user converting two 1080p videos (first video: MOV; second video: MP4; total length: 84 seconds) to 480p resolution. TouchXPRT measures the elapsed time for converting each video and reports the sum of all measured times as the scenario score.

  This test is the same as the test in TouchXPRT 2013.

- Create Music Podcast. This scenario represents a user mixing five audio files (30 minutes total length, WMA, 160Kbps), watermarking the mixed file, and then converting it to an MP3 file for sharing (8:15 minutes length, 160Kbps). TouchXPRT measures the elapsed time for applying the audio watermark and converting to MP3 format, and reports it as the scenario score. Create Music Podcast includes the operations of the TouchXPRT 2013 Export Podcast to MP3 test.

  Improvements from TouchXPRT 2013 include the following:
  - o The test now merges multiple tracks of audio to produce a new piece of music.
  - o New visualization makes it clearer what the test is doing.
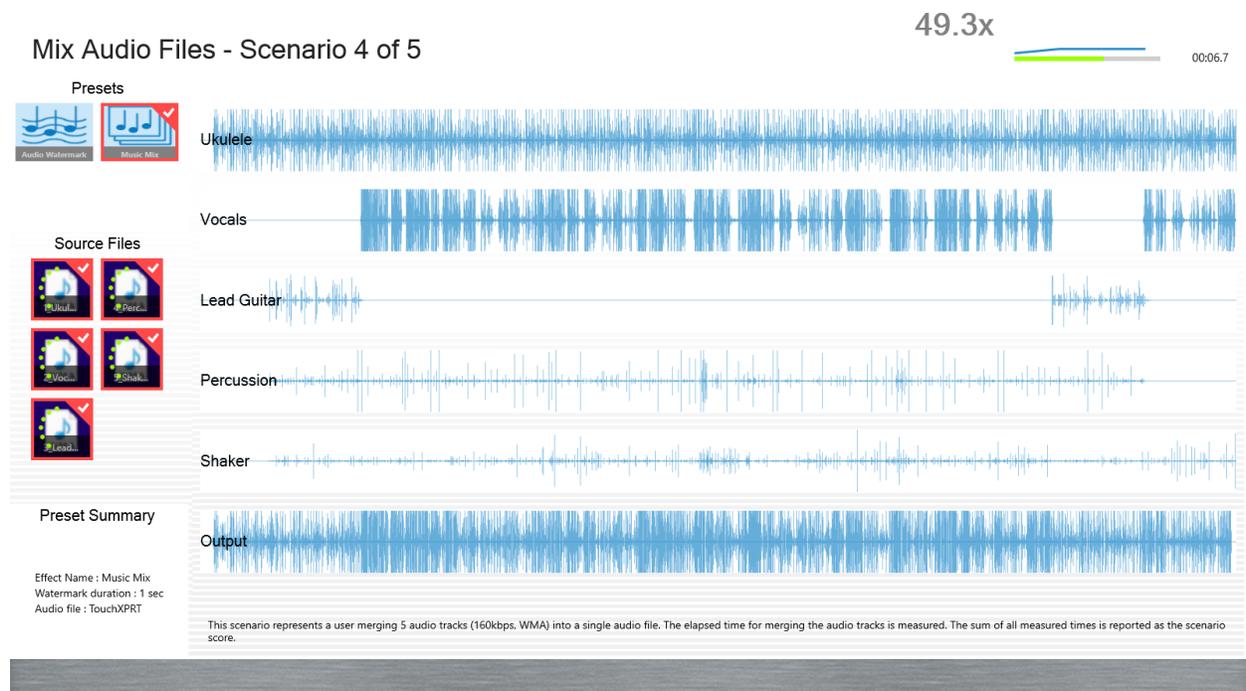 - o The test plays back the resulting music.

Figure 1: Mixing audio files during the Create Music Podcast test.

- **Create Slideshow from Photos.** This scenario represents a user creating a slideshow video (720p, MP4) with photos from an album (24 photos, 16 megapixels). TouchXPRT measures the elapsed time for adding photos to the slideshow and creating a video, and reports the sum of those times as the scenario score.

Improvements from TouchXPRT 2013 include the following:
- o The test is now multithreaded to better reflect the performance of multi-core devices.
- o The photos are higher resolution. They are now 16MP instead of 10.2MP.
- o Testing showed that we could reduce the number of photos without compromising the results. Therefore, we now use 24 photos rather than 50, which shortens the testing time.

SCORING

What do all the test results mean?

TouchXPRT presents a variety of data points about a device's performance during the test. The most important data point, the overall score, is in the upper-right corner of the results page at the end of the test. You can use the overall score to compare your device to others on TouchXPRT.com. However, TouchXPRT provides more than just the overall score; it also gives you information about each individual test.
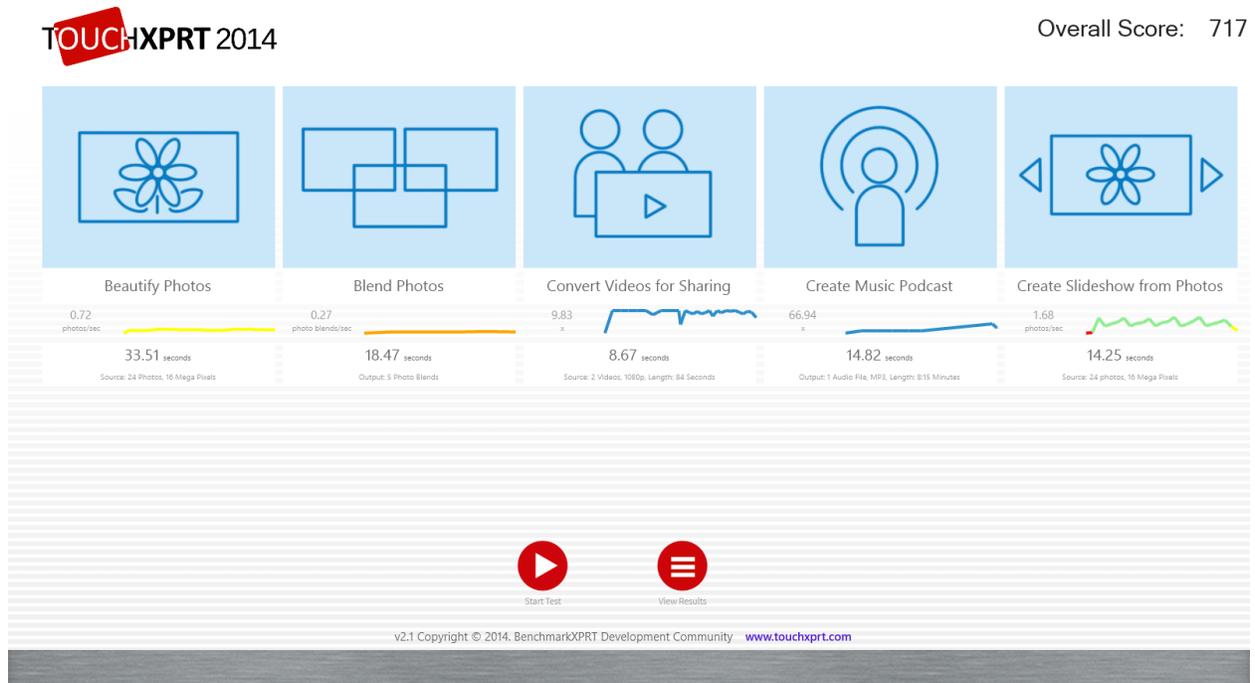

Figure 2: The TouchXPRT 2014 results page.

Here is a description of the data presented on the results page:

- **Colored lines.** Like a lie detector or a seismograph, these lines represent how the device is doing as it goes through its paces. They serve as a graphic indicator of how quickly the device is going at any given point in the test. The colors of the lines indicate the following:
    - o Elapsed time between 0 ms – 150 ms = green (instantaneous)
    - o Elapsed time between 150 ms – 1 s = light green (quick)
    - o Elapsed time between 1 s – 2 s = yellow
    - o Elapsed time between 2 s – 5 s = orange
    - o Elapsed time > 5 s = red.
    - o Blue is a neutral color for longer scenarios where we do not measure responsiveness.
- **Photos/sec.** The photos/sec rate tells you how many photos the device processed per second during a task. The same description applies to the photo blends/sec rate. Higher is better for this kind of score.
- **#x.** This is the ratio of the length of the song or video to the time the test took to run. This measure allows us to compare clips of any length. Higher is better for this type of score as well.
- **Seconds.** This is the time TouchXPRT spent working on that test. Because finishing a task in less time is preferable, a lower number of seconds is better. If you watch the clock, you will notice that this time is less than the time the test took to complete. This is because TouchXPRT realistically implements the way people work, including the way people pause during their work. However, the pauses are the same on all systems, so they do not affect the final score. For this reason, TouchXPRT excludes them from the total test time.
- **Source and Output.** These TouchXPRT constants indicate the quantity and quality of the audio and visual media the test uses. They will be the same on all devices. If your device were a car, it would be like saying that the mileage test took place on a 40-mile track.

Example calculation of the overall score

In this section, we will use the results from a Microsoft Surface Pro 2. As we discussed above, TouchXPRT 2014 comprises five tests. For each test, TouchXPRT records the time the test required to do its work. For the example run on the Surface Pro 2, the times were as follows:

Test name	Time (seconds)
Beautify Photos	33.51
Blend Photos	18.47
Convert Videos for Sharing	8.67
Create Music Podcast	14.82
Create Slideshow from Photos	14.25

The timings only include the time spent working. Wait times are excluded, as are items that take a fixed time, such as playing the music at the end of the Create Music Podcast test. So, the timings reported by TouchXPRT will not typically match the elapsed wall time.

From the timings, TouchXPRT 2014 will compute per-test rates and the overall score. While lower times are better, the per-test rates allow you to compare the tests with a "bigger is better" measure. The table below shows the rates for each test, along with a brief explanation of how they're calculated.

Test name	Rate	How the rate is computed
Beautify Photos (photos/sec)	0.85	Number of photos (24) divided by time.
Blend Photos (photo blends/sec)	0.33	Number of resulting photo blends (5) divided by time.
Convert Videos for Sharing	11.18	Time required to play the video divided by the time to convert the video. The result is how many times longer it would take to play the video than it took to do the conversion.
Create Music Podcast	79.43	Time required to mix the music divided by the time required to play the music. The result is how many times longer it would take to play the music than it took to mix the music.
Create Slideshow from Photos (photos/sec)	2.05	Number of photos (24) divided by time.

To compute the Overall score, we need to scale the timings using the values from a calibration device. For TouchXPRT 2014, the calibration device is a Microsoft Surface RT. The table below shows this.

Test name	Time (seconds)	Calibration time (seconds)	Calibration time divided by time
Beautify Photos	33.51	189.00	5.64
Blend Photos	18.47	226.50	12.26
Convert Videos for Sharing	8.67	60.00	6.92
Create Music Podcast	14.82	95.50	6.44
Create Slideshow from Photos	14.25	88.00	6.18

To compute the Overall score, we take the geometric mean[1] of the resulting scaled values and multiply by 100. The geometric mean of 5.64, 12.26, 6.92, 6.44, and 6.18 is approximately 7.17. Multiplying by 100 gives 717, which is the Overall score reported by TouchXPRT 2014.

AFTER RUNNING THE BENCHMARK

Comparing results to the database

You can view results for TouchXPRT here. To find detailed information on any set of scores, click the link under the Source column. As explained in the help file, results are stored locally in Libraries\Pictures\TouchXPRT2014\Docs\TouchXPRTResult.xml

Submitting results

Click the View Results button to see the results of past runs. To submit a result, click the Submit button associated with that result. Enter your contact information and the amount of RAM in the system, and click Submit.

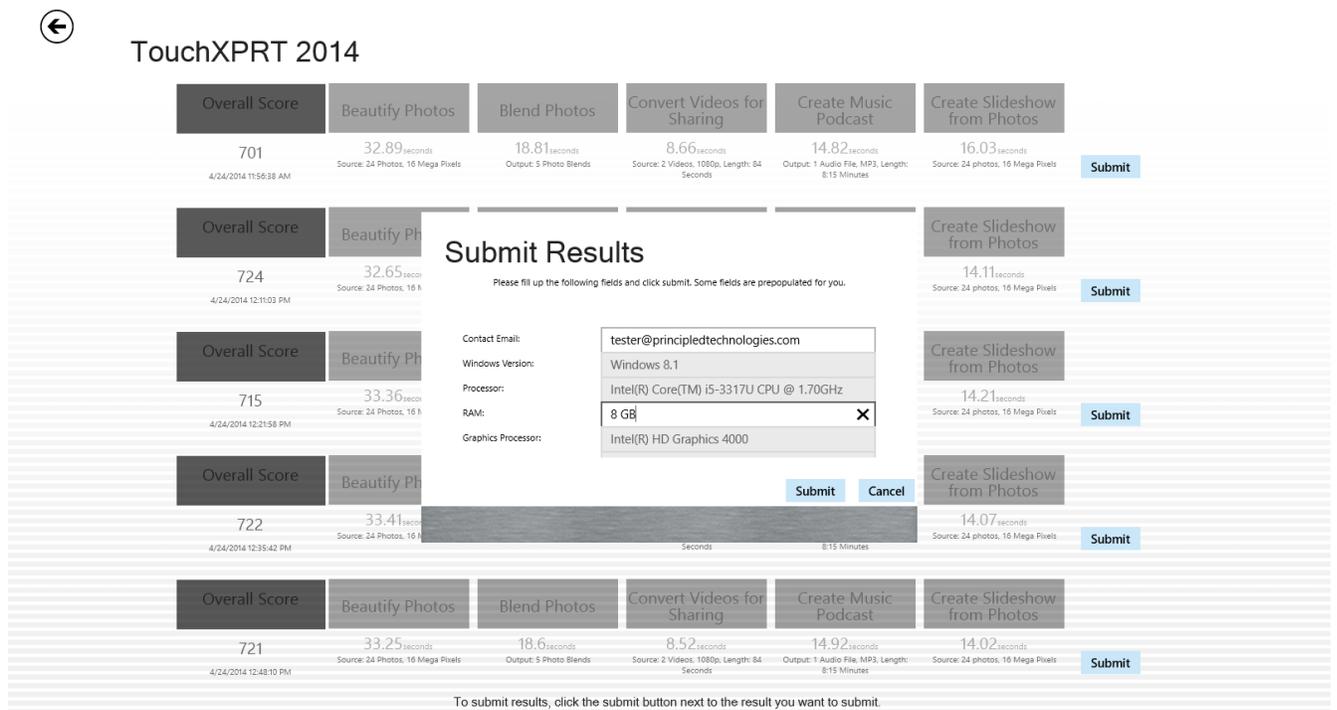[1] en.wikipedia.org/wiki/Geometric_mean

Figure 3: The Submit Results form.

AUTOMATION

TouchXPRT 2014 allows you to automate running your tests. You can run the benchmark from the command line or a script using the start command. The command line options include:

- Start running the test automatically on launch. You choose to run all scenarios and generate an overall score, or run any given scenario.
- Set the number of iterations the test should run.
- Change the delay before the test starts running after launch. By default, TouchXPRT 2014 waits for 5,000 ms before it starts running. This allows the app enough time to load and allows the system to stabilize.
- Change the output filename.

Running TouchXPRT from the command line or a script

There are two ways to launch the app with a command:

1. From the command prompt, launch the app using `start touchxprt2014://`
2. From the Run window (Start+R keys), launch the app using `touchxprt2014://`

The two commands above will take you to the UI. You can also use the following options to run tests without human intervention. Remember that, from a batch file or command prompt, these commands would begin with the start command.

---

1. **To run tests, use the scenario option**

   To run all scenarios and generate an overall score: `touchxprt2014://options/scenario=all`

   To run any particular scenario, use the numbers 1-5, as in this example:
   `touchxprt2014://options/scenario=1`

   The scenarios are numbered as follows:

   1. Beautify Photos
   2. Blend Photos
   3. Convert Videos for Sharing
   4. Create Music Podcast
   5. Create Slideshow from Photos

   Note: The scenario option must be specified each time you launch TouchXPRT from the command line or a script. In the following examples, the scenario switch is set to all, but it can be set to the number of any given scenario.

2. **To specify the number of iterations**

   To specify the number of iterations: `touchxprt2014://options/scenario=all/iterations=2`

3. **To change the default five-second delay**

   By default when the app is launched, it waits 5,000 ms before it starts running tests. This delay allows the app to load and the system to stabilize. To change the start delay, use the startdelay option to specify the delay in milliseconds. This example shows how to change the delay to 6,000 ms:
   `touchxprt2014://options/scenario=all/startdelay=6000`

4. **To change the output filename**

   All results will save to Pictures\TouchXPRT2014\Docs. By default, the results are stored in a file named results.xml. However, you can change the name of the output file to identify the test according to your conventions.

   For example, to change the filename to MyTestSystem.xml:
   `touchxprt2014://options/scenario=all/filename=MyTestSystem.xml`

These commands can be combined:
`touchxprt2014://options/scenario=all/iterations=3/startdelay=6000/filename=MyTestSystem.xml`

General notes about scripting

All switches are optional. As noted above, however, a result of this behavior is that omitting the scenario option causes TouchXPRT 2014 to pause at its UI. As a general rule, we do not recommend mixing scripting and manual runs, so we encourage the use of the scenario option.

TouchXPRT 2014 ignores invalid switches or switches with invalid values. In those cases, testing proceeds using TouchXPRT's typical defaults.

During the run, TouchXPRT 2014 will store its results in a temporary file. Scripts should test for the existence of the results file to determine that the test is complete. There are several ways to do this. We offer an example of how to

test from a DOS batch file below. This example polls every 5 seconds and assumes you are using the default result file name of results.xml.

```
start touchxprt2014://options/scenario=all
:WAITRESULT
IF EXIST %homepath%\pictures\touchxprt\docs\results.xml GOTO DONE
    timeout /t 5 /nobreak > nul
goto WAITRESULT
:DONE
```

Proceed with the testing logic

ABOUT THE BENCHMARKXPRT BENCHMARKS

The BenchmarkXPRT tools are a set of apps that help you test how well devices do the kinds of things you do every day. In addition to TouchXPRT 2014, the BenchmarkXPRT suite currently comprises the following tools:

- BatteryXPRT for Android, an app to estimate the battery life of Android devices quickly and reliably
- MobileXPRT, an app to test the responsiveness of Android devices
- WebXPRT, an online tool to test the Web browsing capabilities of any device with Internet access
- HDXPRT, a program that uses commercial applications to test the capabilities and responsiveness of PCs

We designed the apps to test a wide range of devices on a level playing field. When you look at results from XPRTs, you get unbiased, fair product comparison information.

The community model

We built BenchmarkXPRT around a unique community model. Community membership is open to anyone, and there are many different ways to participate. Members of the BenchmarkXPRT Development Community are involved in every step of the process. They give input on the design of upcoming benchmarks, contribute source code, and help test the benchmarks. Community members have access to the source code and access to early releases of the benchmarks in the form of community previews.

The community helps us avoid the ivory tower syndrome. Diversity of input during the design process makes the tests more representative of real world activity. Giving community members access to the source code both improves the implementation of the benchmark and increases confidence in the code.

The primary difference between the community model and an open source model is the control of derivative works. It is important that the BenchmarkXPRT benchmarks return consistent results. If the testing community calls different derivative works by the same name, the result would be test results that are not comparable. That would limit, if not destroy, the tools' effectiveness.

Where do I get more information?

Go to TouchXPRT.com or follow us on Twitter and Facebook. We announce breaking news on the BenchmarkXPRT blog (available to everyone) and the BenchmarkXPRT forums (available to members only). If you cannot find the answer to your question, or you need help with TouchXPRT, send an email to our team at BenchmarkXPRTsupport@principledtechnologies.com.

What is the BenchmarkXPRT Development Community?

The BenchmarkXPRT Development Community is a forum where registered members can contribute to the process of creating and improving benchmarks, including TouchXPRT. If you are not currently a community member, we encourage you to join! (Yes, that means you – our community is open to everyone, from software developers to interested consumers.) Not only will you get early releases of future versions of TouchXPRT, but you will also be able to download the source code (available to members only) and influence the future of the app. Register now, or for more information, see the BenchmarkXPRT FAQ.

CONCLUSION

We hope this paper has answered any questions you may have about TouchXPRT. If you have suggestions about ways to improve the benchmark, or if you have any other questions, please post them on the community forum or e-mail us at BenchmarkXPRTsupport@principledtechnologies.com. For more information, visit us at www.benchmarkxprt.com and www.touchxprt.com.

ABOUT PRINCIPLED TECHNOLOGIES

Principled Technologies, Inc.
1007 Slater Road, Suite 300
Durham, NC, 27703
www.principledtechnologies.com

We provide industry-leading technology assessment and fact-based marketing services. We bring to every assignment extensive experience with and expertise in all aspects of technology testing and analysis, from researching new technologies, to developing new methodologies, to testing with existing and new tools.

When the assessment is complete, we know how to present the results to a broad range of target audiences. We provide our clients with the materials they need, from market-focused data to use in their own collateral to custom sales aids, such as test reports, performance assessments, and white papers. Every document reflects the results of our trusted independent analysis.

We provide customized services that focus on our clients' individual requirements. Whether the technology involves hardware, software, Web sites, or services, we offer the experience, expertise, and tools to help our clients assess how it will fare against its competition, its performance, its market readiness, and its quality and reliability.

Our founders, Mark L. Van Name and Bill Catchings, have worked together in technology assessment for over 20 years. As journalists, they published over a thousand articles on a wide array of technology subjects. They created and led the Ziff-Davis Benchmark Operation, which developed such industry-standard benchmarks as Ziff Davis Media's Winstone and WebBench. They founded and led eTesting Labs, and after the acquisition of that company by Lionbridge Technologies were the head and CTO of VeriTest.

Disclaimer of Warranties; Limitation of Liability:

PRINCIPLED TECHNOLOGIES, INC. HAS MADE REASONABLE EFFORTS TO ENSURE THE ACCURACY AND VALIDITY OF ITS TESTING, HOWEVER, PRINCIPLED TECHNOLOGIES, INC. SPECIFICALLY DISCLAIMS ANY WARRANTY, EXPRESSED OR IMPLIED, RELATING TO THE TEST RESULTS AND ANALYSIS, THEIR ACCURACY, COMPLETENESS OR QUALITY, INCLUDING ANY IMPLIED WARRANTY OF FITNESS FOR ANY PARTICULAR PURPOSE. ALL PERSONS OR ENTITIES RELYING ON THE RESULTS OF ANY TESTING DO SO AT THEIR OWN RISK, AND AGREE THAT PRINCIPLED TECHNOLOGIES, INC., ITS EMPLOYEES AND ITS SUBCONTRACTORS SHALL HAVE NO LIABILITY WHATSOEVER FROM ANY CLAIM OF LOSS OR DAMAGE ON ACCOUNT OF ANY ALLEGED ERROR OR DEFECT IN ANY TESTING PROCEDURE OR RESULT.

IN NO EVENT SHALL PRINCIPLED TECHNOLOGIES, INC. BE LIABLE FOR INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH ITS TESTING, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. IN NO EVENT SHALL PRINCIPLED TECHNOLOGIES, INC.'S LIABILITY, INCLUDING FOR DIRECT DAMAGES, EXCEED THE AMOUNTS PAID IN CONNECTION WITH PRINCIPLED TECHNOLOGIES, INC.'S TESTING. CUSTOMER'S SOLE AND EXCLUSIVE REMEDIES ARE AS SET FORTH HEREIN.